Optimal Interval Partitioning for Temporal Databases

Thomas Zurek

University of Edinburgh

Overview

- temporal databases
- partitioned equi-joins
- partitioned temporal joins
- interval partitioning problem
- algorithm `IP-opt`

Temporal Relations

Relation **Hamlet**

City	Timestamp
Berlin	[3,8]
Dublin	[2,10]
London	[4,9]
Madrid	[1,8]
Oslo	[6,10]
Vienna	[1,10]

Relation **Faust**

City	Timestamp
Bern	[1,4]
London	[1,4]
Munich	[7,9]
Oslo	[2,5]
Paris	[1,3]
Rome	[7,10]

Hamlet $\bowtie_C$ Faust

with $C \equiv$ Hamlet.City = Faust.City

$$\downarrow$$

Hamlet$_1$ $\bowtie_C$ Faust$_1$

Hamlet		Faust	
Berlin	[3,8]	Bern	[1,4]
Dublin	[2,10]	London	[1,4]
London	[4,9]		

A – L

Hamlet$_2$ $\bowtie_C$ Faust$_2$

Hamlet		Faust	
Madrid	[1,8]	Munich	[7,9]
Oslo	[6,10]	Oslo	[2,5]

M – O

Hamlet$_3$ $\bowtie_C$ Faust$_3$

Hamlet		Faust	
Vienna	[1,10]	Paris	[1,3]
		Rome	[7,10]

P – Z

Hamlet $\bowtie_C$ Faust

with $C \equiv$ Hamlet.Timestamp $intersects$ Faust.Timestamp

$$\downarrow$$

Hamlet$_1$ $\bowtie_C$ Faust$_1$

Hamlet		Faust	
Berlin	[3,8]	Bern	[1,4]
Dublin	[2,10]	London	[1,4]
Madrid	[1,8]	Oslo	[2,5]
Vienna	[1,10]	Paris	[1,3]

[1,3]

Hamlet$_2$ $\bowtie_C$ Faust$_2$

Hamlet		Faust	
Berlin	[3,8]	Bern	[1,4]
Dublin	[2,10]	London	[1,4]
London	[4,9]	Oslo	[2,5]
Madrid	[1,8]		
Vienna	[1,10]		

[4,5]

Hamlet$_3$ $\bowtie_C$ Faust$_3$

Hamlet		Faust	
Berlin	[3,8]	Munich	[7,9]
Dublin	[2,10]	Rome	[7,10]
London	[4,9]		
Madrid	[1,8]		
Oslo	[6,10]		
Vienna	[1,10]		

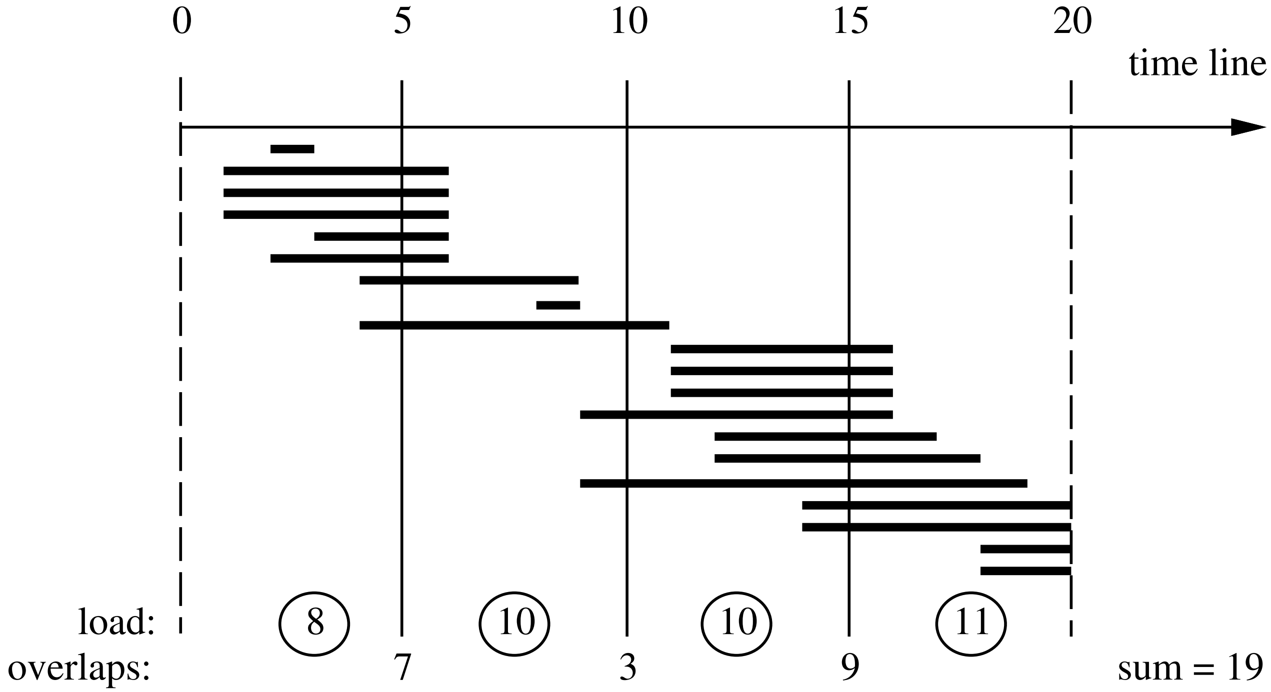
[6,10]

Partitioning is relevant for . . .

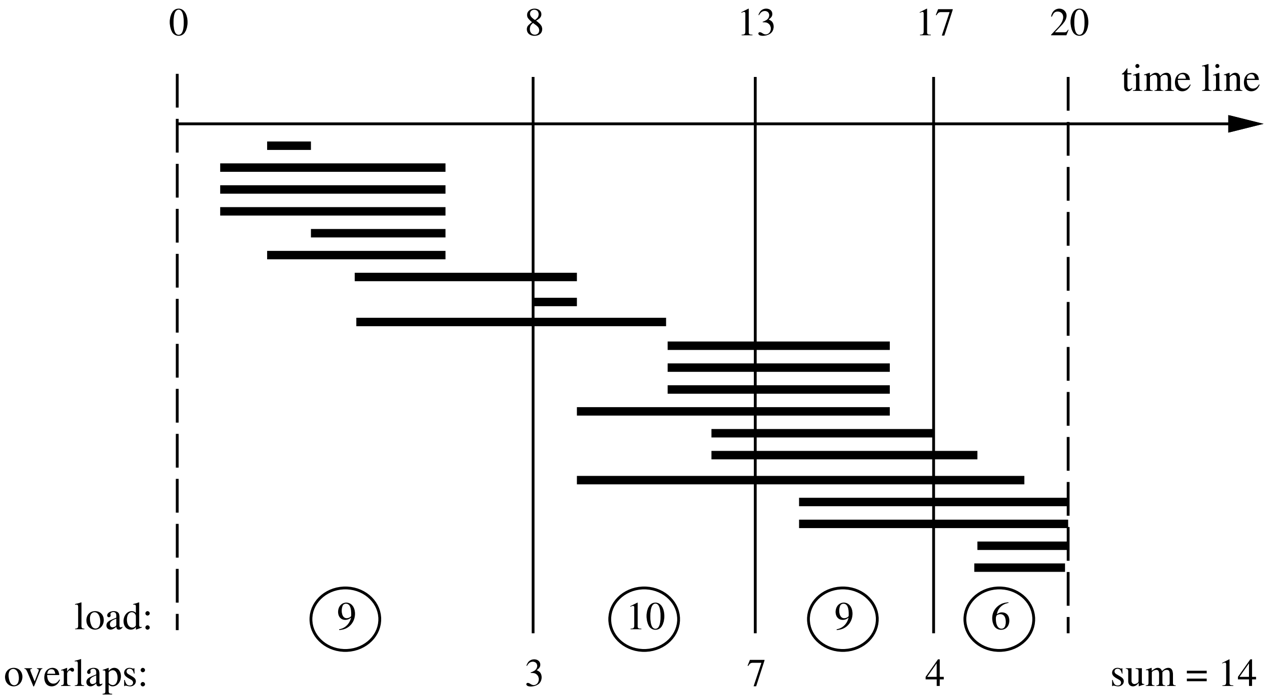
- hash join processing
- parallel join processing
- temporal index trees
- physical data partitioning

All these areas can benefit from a reduction of redundancy.

An example scenario



and another partition ...



Interval Partitioning Problem (IP)

Instance

- A collection U of intervals $\langle u_1, \ldots, u_N \rangle$,

- a positive number X.

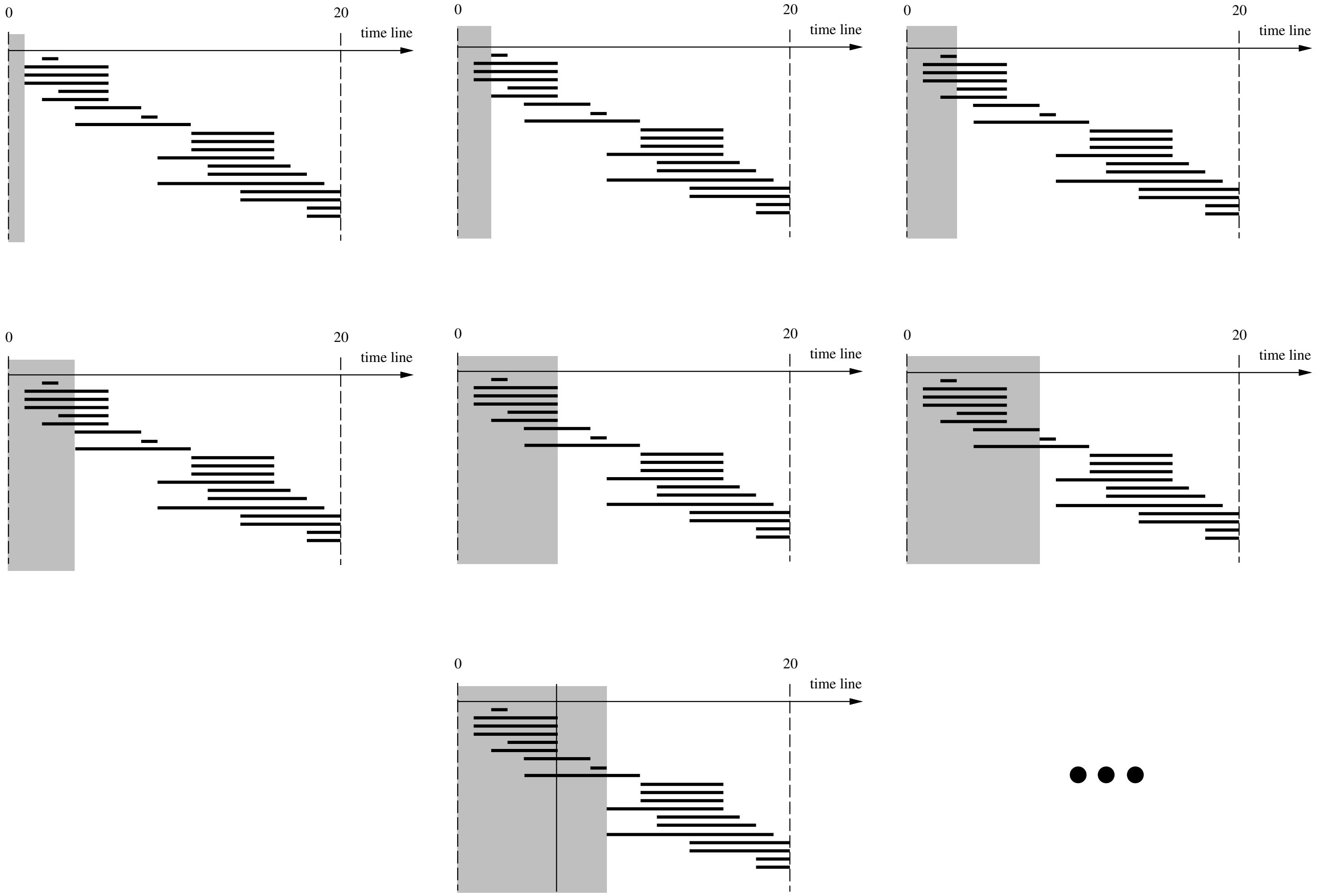
Question

Is there a partition $P = \{p_1, \ldots, p_{m-1}\}$ of U that minimises

$$\sum_{p \in P} \mathit{overlaps}(p)$$

such that $\quad \mathit{load}(p_{j-1}, p_j) \leq X \quad$ for $j = 1, \ldots, m$?

(Rem.: $p_0 = -\infty$, $\quad p_m = +\infty$)

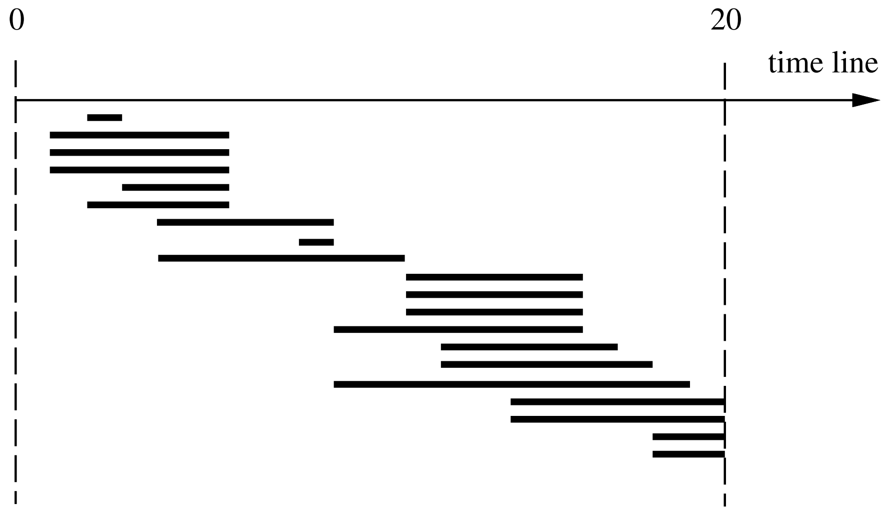

0
20
time line
0
20
time line
0
20
time line
0
20
time line
0
20
time line
0
20
time line
0
20
time line

Algorithmic step at time t

1. $J \;=\; \{$those $t' < t$ such that $load(t', t) \leq X\}$

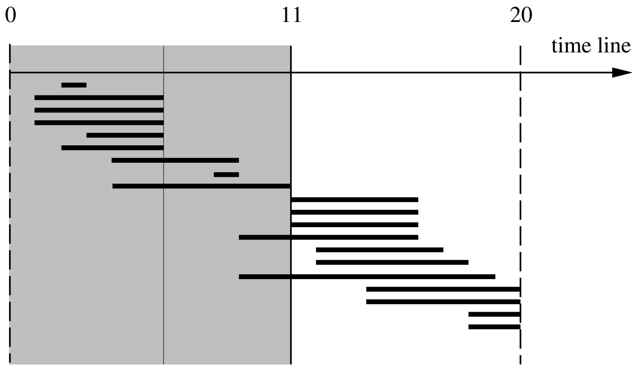
2. $pred(t) \;=\;$ the $t' \in J$ that minimises $overlaps(t') + c(t')$

3. $c(t) \;=\; overlaps(pred(t)) \;+\; c(pred(t))$



i	q_i	$overlaps(q_i)$	J (for $X = 10$)	$pred(q_i)$	$c(q_i)$
0	0	0	—	—	0

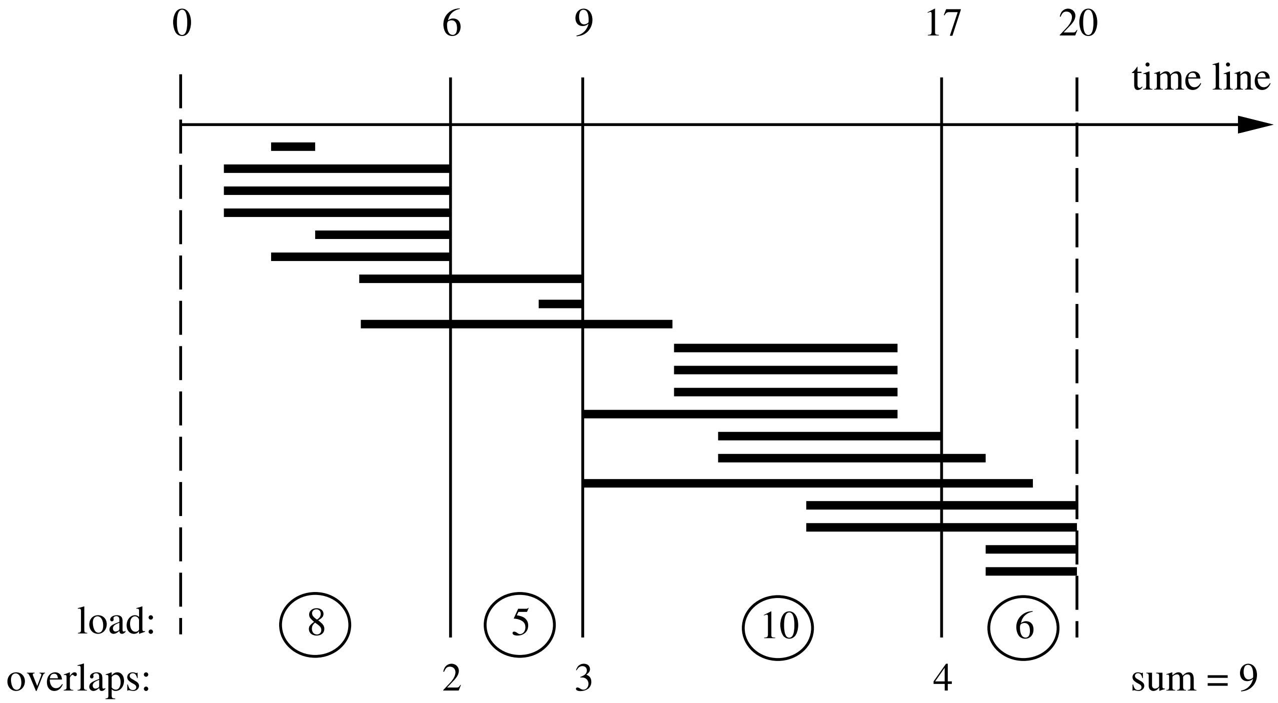
i	q_i	$overlaps\,(q_i)$	J (for $X = 10$)	$pred\,(q_i)$	$c(q_i)$
0	0	0	—	—	0
1	1	3	$\{q_0\}$	$q_0 = 0$	0



i	q_i	$overlaps\,(q_i)$	J (for $X = 10$)	$pred\,(q_i)$	$c(q_i)$
0	0	0	—	—	0
1	1	3	$\{q_0\}$	$q_0 = 0$	0
2	2	5	$\{q_0, q_1\}$	$q_0 = 0$	0
3	3	5	$\{q_0, q_1, q_2\}$	$q_0 = 0$	0
4	4	7	$\{q_0, q_1, q_2, q_3\}$	$q_0 = 0$	0
5	6	2	$\{q_0, q_1, q_2, q_3, q_4\}$	$q_0 = 0$	0
6	8	2	$\{q_0, q_1, q_2, q_3, q_4, q_5\}$	$q_0 = 0$	0
7	9	3	$\{q_3, q_4, q_5, q_6\}$	$q_5 = 6$	2
8	11	5	$\{q_5, q_6, q_7\}$	$q_5 = 6$	2

i	q_i	$overlaps\,(q_i)$	J (for $X = 10$)	$pred\,(q_i)$	$c(q_i)$
0	0	0	—	—	0
1	1	3	$\{q_0\}$	$q_0 = 0$	0
2	2	5	$\{q_0, q_1\}$	$q_0 = 0$	0
3	3	5	$\{q_0, q_1, q_2\}$	$q_0 = 0$	0
4	4	7	$\{q_0, q_1, q_2, q_3\}$	$q_0 = 0$	0
5	**6**	**2**	$\{\mathbf{q_0}, \mathbf{q_1}, \mathbf{q_2}, \mathbf{q_3}, \mathbf{q_4}\}$	$\mathbf{q_0 = 0}$	**0**
6	8	3	$\{q_0, q_1, q_2, q_3, q_4, q_5\}$	$q_0 = 0$	0
7	**9**	**3**	$\{\mathbf{q_3}, \mathbf{q_4}, \mathbf{q_5}, \mathbf{q_6}\}$	$\mathbf{q_5 = 6}$	**2**
8	11	5	$\{q_5, q_6, q_7\}$	$q_5 = 6$	2
9	12	7	$\{q_5, q_6, q_7, q_8\}$	$q_5 = 6$	2
10	14	9	$\{q_7, q_8, q_9\}$	$q_7 = 9$	5
11	16	5	$\{q_7, q_8, q_9, q_{10}\}$	$q_7 = 9$	5
12	**17**	**4**	$\{\mathbf{q_7}, \mathbf{q_8}, \mathbf{q_9}, \mathbf{q_{10}}, \mathbf{q_{11}}\}$	$\mathbf{q_7 = 9}$	**5**
13	18	5	$\{q_{11}, q_{12}\}$	$q_{12} = 17$	9
14	19	4	$\{q_{11}, q_{12}, q_{13}\}$	$q_{12} = 17$	9
15	20	0	$\{q_{11}, q_{12}, q_{13}, q_{14}\}$	$q_{12} = 17$	9

The Result



Summary

- temporal relations + relevance of partitioning
- interval partitioning problem
- `IP-opt`:

  dynamic programming
  $O(n^2)$

- possible applications:
  temporal joins, temporal index structures, physical partitioning